

*Shaping the Library of the Future:
Digital Library Developments at
Los Alamos National Laboratory's
Research Library*

Los Alamos
NATIONAL LABORATORY

*Los Alamos National Laboratory is operated by the University of California
for the United States Department of Energy under contract W-7405-ENG-36.*

An Affirmative Action/Equal Opportunity Employer

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither The Regents of the University of California, the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by The Regents of the University of California, the United States Government, or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of The Regents of the University of California, the United States Government, or any agency thereof.

Issued: October 1994

*Shaping the Library of the Future:
Digital Library Developments at Los
Alamos National Laboratory's Research Library*

Richard E. Luce

Shaping the Library of the Future: Digital Library Developments at Los Alamos National Laboratory's Research Library

by
Richard E. Luce

(Conference paper presented at INFOTECH '94, October 25-26, 1994, Oak Ridge, TN)

Abstract

This paper offers an overview of current efforts at the Research Library, Los Alamos National Laboratory, (LANL), to develop digital library services. Current projects of LANL's Library without Walls initiative are described. Although the architecture of digital libraries generally is experimental, subject to debate, and evolving, one principle of LANL's approach to delivering library information is the use of Mosaic as a client for the Research Library's resources. Several projects under development have significant ramifications for delivering library services over the Internet. Specific efforts via Mosaic include support for preprint databases, providing access to citation databases, and access to a digital image database of unclassified Los Alamos technical reports.

Introduction

During the last several years, the conceptual paradigm of special research libraries has changed from a focus on the building which houses the physical collections to that of an information access service, bounded neither by the walls of the library building, nor by traditional book and journal collections. In this paradigm, patrons connect remotely and use technology to access local and worldwide library, public, and commercial information providers -- a concept known as the **digital library**.

A variety of factors have propelled virtual library developments to the forefront of information science, including: rapidly advancing information technologies coupled with customer expectations that libraries explore every avenue to deliver information effectively and efficiently; plans for a National Information Infrastructure, with its associated requirements for storing and accessing vast amounts of digital data over the information superhighway; and limited financial resources for the acquisition of ever-growing scientific publications, which continue to outpace inflation.

In late 1993, Los Alamos National Laboratory (LANL) underwent a massive reorganization. The Research Library was placed within the new Computing, Information,

and Communications Division (CIC) of the Laboratory. The reorganization provided the Library with an opportunity to start, or to raise the visibility of, several major strategic initiatives focusing on information management. One of the most significant is the *Library without Walls*.

The Vision

The LANL Research Library vision for library services seeks to combine the best information resources, staff, and technology to deliver world-class service to our research community. It embodies the belief that we will become a leader in providing access to global electronic information resources by anticipating and meeting customer needs. One manifestation of those needs, expressed or implied, is the delivery of information to the researchers' desktops -- wherever and whenever they need it -- from digital library resources. This is the essence of the *Library without Walls*.

Several key goals and principles define the *Library without Walls* initiative:

- The virtual library is not a single entity at Los Alamos but requires the seamless integration of other virtual library resources through technology linkages;
- Worldwide access to the digital resources of the Research Library and the Laboratory's scientific research;
- The convergence of many of distinct efforts, encompassing a variety of fields will be required;
- New multi-media, inter-active compound documents and digital artifacts that extend beyond the linear capabilities represented by print publications must be incorporated;
- Systems and products must be measured by their ability to enhance new forms of collaboration among our users.

The long term goal is the creation of a network of knowledge systems and machines which facilitate synergy and collaboration between people.

Laying the Foundation

Before embarking on a digital library initiative, a sound infrastructure which improved our information technology foundation and provided appropriate tools was necessary in order to build our service capabilities. The following three capabilities were integrated to provide that foundation:

(1) Upgrading the Online Catalog. In 1992, the Library upgraded its automated library system and online catalog by migrating to the Geac Advance system. Previously, the automated Geac GLIS system had been viewed as the centerpiece of library automation and technology efforts. That notion was replaced with a conceptual framework in which the new automated system became one of several modules, each of which had to be integrated to meet a multiplicity of information access and delivery requirements.

External access to the library's online catalog was now viewed as a prerequisite for creating resource sharing partnerships with other libraries and opening up greater public access.

Migration to the Geac Advance system not only achieved these goals but offered several other significant advantages: cost-effective migration to an open hardware and software platform; substantially larger capacity to support simultaneous users; control over data migration; easy connection of the online catalog to the Internet; and the ability to load external citation databases. The online catalog became publicly available through either telnet, Gopher, or Mosaic in VT 100 mode in 1993. Later installation of a smaller Advance system will allow similar support for classified technical reports.

(2) Modernizing the Workstation Toolset. Providing an appropriate technology toolset for Library staff was imperative to enhancing our collective technology knowledge and ability to respond to user requests. For Library staff this meant migrating from common area dumb terminals and a few stand-alone personal computers to workstations that had sufficient capabilities to access developing library modules. Three basic workstation goals were established: all support staff would have appropriate workstations on their desks; workstations on the desktop must be able to access Library systems; and workstations had to be connected to internal and external networks.

(3) Developing a Robust Local Area Network (LAN). The key criteria for developing the LAN were integration of the desktop tools with CD-ROM databases, access to Internet resources, shared application software, and electronic mail to facilitate communication. A Novell LAN was installed in 1992 to meet these criteria.

That network has grown to three Novell servers supporting 65 rack-mounted networked CD readers, 20 GB of RAID-5 magnetic storage for CD applications to optimize performance, and three Kubix jukeboxes (240 CD's each) to deliver images from UMI's ABI Inform. We are beginning to provide support for external (campus) connections via 12 LogiCraft OmniWare devices ('486/50 PC's) which allow outside users to access the Library's CD-ROM systems. Access is provided through an X-window client providing the full capabilities allowed by the CD application in native PC mode.

Current Projects

The heterogeneous computing environment at LANL places a premium on the ability to deliver services across multiple platforms via TCP/IP. The exponential growth of the WorldWideWeb, with Mosaic as a preferred client tool, has had significant implications for library access and project application developments. Consequently, ubiquitous access to Mosaic is driving several efforts to provide library information through Mosaic to the desktop.

Constructing the *Library without Walls* involves several interrelated projects that are at varying stages of completion, each solving different facets of the digital library goals.

Preprints: The High-Energy Physics Preprint Service

A science library's ability to provide access to preprint literature is strategic to address the issue of timely communication of scholarly research. A *Library without Walls* is now a growing reality for the distribution of scientific preprints.

In August 1991, in response to the needs of the particle-theory community, Paul Ginsparg of the Laboratory's Theoretical Division created a fully automated hybrid electronic bulletin board system for preprints in high-energy physics (xxx.lanl.gov/). Papers (typically preprints) are submitted electronically to the system, which assigns reference numbers, and makes them available to subscribers via a wide variety of conventional means of network access. The system not only provides titles, which *Spires* at SLAC has supported for years, but also provides entire papers rapidly. Users can submit preprints, read abstracts, and then request a complete preprint, which is transmitted in a few seconds, all free of charge.

The simplest means of access is electronic mail, through which remote users may get help on available commands, obtain listings for given periods, search the listings for author names or keywords in titles, and so on. Subscribers to the system receive a daily listing of titles and abstracts of new submissions received. The original author of a paper can incorporate ongoing corrections and addenda. The system also makes available anonymous FTP access to the papers and the listings directories, and offers access by other network utilities such as WAIS, Gopher, and the WorldWideWeb.

In the first six months the service had 2,000 users, mostly in formal quantum field theory and string theory. A year later, as Ginsparg expanded his network to other sub-disciplines of physics, some 8,000 subscribers were submitting nearly 600 preprints each month. As of summer 1994, there are over 22,000 users worldwide, accounting for 35,000 daily transactions, and 20,000 preprint papers are stored on the system, at a cost of about 5 cents per paper plus network transmission costs.

The key to the system's popularity is its convenience. Access requires no special software, and storage and retrieval of figures do not pose problems. Multiple figure files are typically submitted as compressed Postscript, and papers can be printed with the figures embedded directly in the text. The system has been improved with the addition of an automatic archiving function and added search functions that allow the user to select any other preprint referenced by the one he is reading. A cross-referencing feature is available to permit linkages among bulletin boards in different fields. Many researchers consider this hybrid bulletin board system an essential tool and it is particularly useful in areas of the world without libraries.

Over the past three years, the software has been thoroughly tested and debugged, and now represents an automated system highly robust to the vagaries---including computer crashes, misconfigured return addresses, incorrectly formatted text files, oversized graphics files---of electronic networking and large numbers of users. The present configuration at Los Alamos uses a combination of e-mail, anonymous FTP, and window-oriented utilities, such as Gopher and the WorldWideWeb combined with WAIS indexing. Recently, macros have been developed to grab files from the e-print archives and hypertext them. The macros are set up to automatically generate only linkages internal to files. However, external URL's can easily be added by the user, others could be added automatically.

The responsibility for software maintenance and development has been transferred to computer staff under the direction of the Research Library, which is interested in both

scientific communications archiving and electronic publishing. As yet unanswered questions about the future of the system include technical issues, such as what network utilities are best suited for archiving and retrieving information, and social issues such as questions about the need for a review process for submitted papers and whether new forms of peer recognition for authors will be needed. The near future may bring improved technical capabilities, providing interactive features, with users solving and graphing equations online, calling up windows with references to select from, and reading comments of reviewers. The social issues will take longer to resolve and the American Physical Society is convening a forum in October 1994 to discuss these issues and their impact on communication of scientific research.

Preprint bulletin boards are currently in operation in nine disciplines: high energy particle theory, particle phenomenology, general relativity and quantum cosmology, nuclear theory, condensed matter theory, algebraic geometry, computational physics, astrophysics, and functional analysis. The operation of preprint systems provides a permanent and indexed archive of electronically accessible information and a mechanism for information exchange and dissemination particularly appropriate in interdisciplinary or emerging areas of scientific research.

Access to Published Literature via Mosaic

Access to published literature in the form of indexes, abstracts and alert services remains vital to keep pace with developing multi-disciplinary research. While the Research Library has invested heavily in CD-ROM products which provide access to indexes, abstracts, and full-text images, mixing a variety of CD products from different providers creates severe limitations attempting to search across multiple years or databases. These limitations, coupled with the high cost of providing external access, impede access and usage in a heterogeneous environment. The challenge is how to get beyond these limitations.

In order to work within our goal of delivering information services via Mosaic as the client interface, the Research Library embarked on a development effort to test the delivery of Science Citation Index (SCI) database via Mosaic, with Verity's Topic serving as the search engine underneath Mosaic. SCI was chosen for the pilot effort due to the comprehensive range of science which it covers, the corresponding appeal to the interdisciplinary research interests of the Laboratory, and the proven value of citation trails.

A prototype of the database with abstracts is now running and we are completing arrangements to license SCI data back through 1980. The pilot successfully demonstrated the ability to parse the data into the format required by Topic and integrate search results in the HTML format required by Mosaic. Of special research interest is the fact that this database grows rapidly (80 MB weekly) and the full set is large enough to require parallel processing to support broad multi-year queries.

Mounting other index and abstract databases is progressing simultaneously. The Research Library is involved in an effort to load the Energy Database (EDB), produced by DOE's Office of Scientific and Technical Information and the index to UMI's ABI Inform, as citation databases using Geac Advance.¹ The EDB is a multidisciplinary database of 2.5 million records containing references to information related to energy research and

¹ Holtkamp, Irma S. *The Energy Science and Technology Database on a Local Library System: A Case Study at the Los Alamos National Laboratory Research Library*. Conference paper presented at INFOTECH '94, October 25-26, 1994, Oak Ridge, TN.

technology from worldwide sources. ABI Inform is a full image database of business periodicals.

The Library is experimenting with a Mosaic interface to its online catalog via a Z39.50 connection to the Advance system. This effort will allow the EDB and ABI Inform databases to be accessible to LANL researchers through Mosaic. As a part of the cooperative development efforts between Geac and Los Alamos, the Z39.50 work could be extended to link to other Z39.50 library servers outside the Laboratory, thereby providing a common interface to heterogeneous external library catalogs.

Image Database of Technical Reports

Users clearly want access to end-products rather than information about the end-products. To meet this expectation we must solve the issues surrounding access to and retrieval of full images.

Library without Walls projects are tackling these issues by providing desktop digital access to unclassified Los Alamos technical report images. The initial goal is to put the full text and scanned page images of over 5,000 Los Alamos technical reports into a locally mounted network-based server with an easy to use retrieval protocol. Unclassified Los Alamos technical reports were a logical beginning step to develop and test these capabilities because they are a part of our institutional memory and no copyright issues exist. These technical reports go back to 1943 and contain reports of research, conference proceedings, and environmental reports. Furthermore, this effort will support our goal of making the products of Los Alamos scientific research widely available to the public.

Current efforts are centered around the LAROLA project (**L**os **A**lamos **R**eports **O**n**L**ine **A**rchives), a collaborative effort involving three groups within CIC Division. LAROLA is based on the principle of integrating public domain software to ensure wide access and facilitate easy software replication. LAROLA uses Mosaic as the client and supports WAIS as the search and retrieval database engine. Some of the more complex research and development issues this project faces include:

- supporting viewers that will work well on virtually any network machine over the Internet;
- navigating through large image and text files;
- authenticating archival digital reports;
- high quality production scanning and OCRing of microfiche and old paper archives;
- finding cost-effective and high performance storage for very large archive image files;
- developing a transparent link between the scanned images and the bibliographic information in the library's online catalog.

A prototype of LAROLA is online and ongoing iteration of the user interface continues to improve the model.

Other Efforts

We are witnessing numerous changes in how we do business as a result of implementing the foundation capabilities, described at the beginning of this paper. While these have not required the kind of development efforts of the various *Library without Walls* projects described herein, they are nevertheless having a major impact on our day-to-day work. A few examples:

- profiled news articles and reference questions are routinely delivered via electronic mail;
- access to a monthly Library newsletter through Mosaic and Gopher;
- support for forms which allow requests via Mosaic and Gopher for reference requests, literature searches, book requests, and technical report delivery;
- listing of What's New in books, training schedules, special reports, and news flashes;
- exploration of Mosaic as a means to deliver *Library without Walls* services to the University of New Mexico's Taos Community College in support of the Laboratory's outreach efforts.

These efforts have provided a platform for us to learn and explore new avenues of delivering information to our customers.

Summary

This paper has outlined LANL's Research Library's approach to developing integrative information technologies with broad access capabilities to provide virtual library services to its customers. To solve the questions inherent in this approach will involve drawing on the disciplines of computer science, information technology, information science, and sociology. The common theme of all the *Library without Walls* projects described in this paper is the use of the latest non-proprietary client tools preferred by the research community (e.g., Mosaic) which are accessible over the Internet. The Research Library has taken significant steps to not only deliver digital information to the Laboratory, but also to the worldwide scientific community and the general public via the Internet.

Richard E. Luce is the Group Leader of the Research Library and Project Leader of the *Library Without Walls* for the Los Alamos National Laboratory. He has an extensive background and expertise integrating information technology into libraries and managing library automation efforts.

